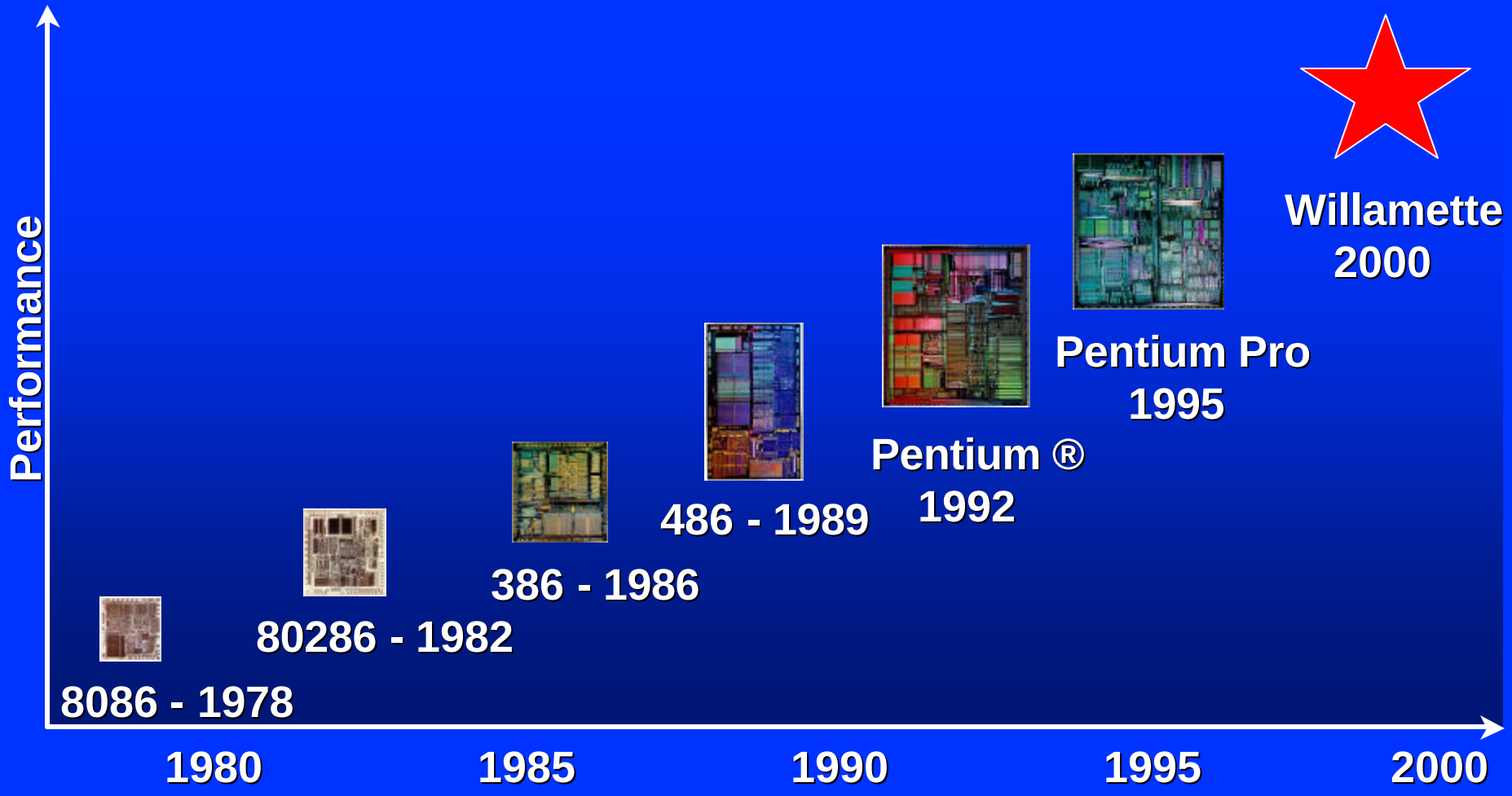
Willamette: Next Generation IA-32 Micro-architecture

Glenn Hinton
Intel Fellow
Performance Microprocessor Division
Intel Corporation

February 15, 2000

Intel Developer Forum Spring 2000

Intel Desktop Core History



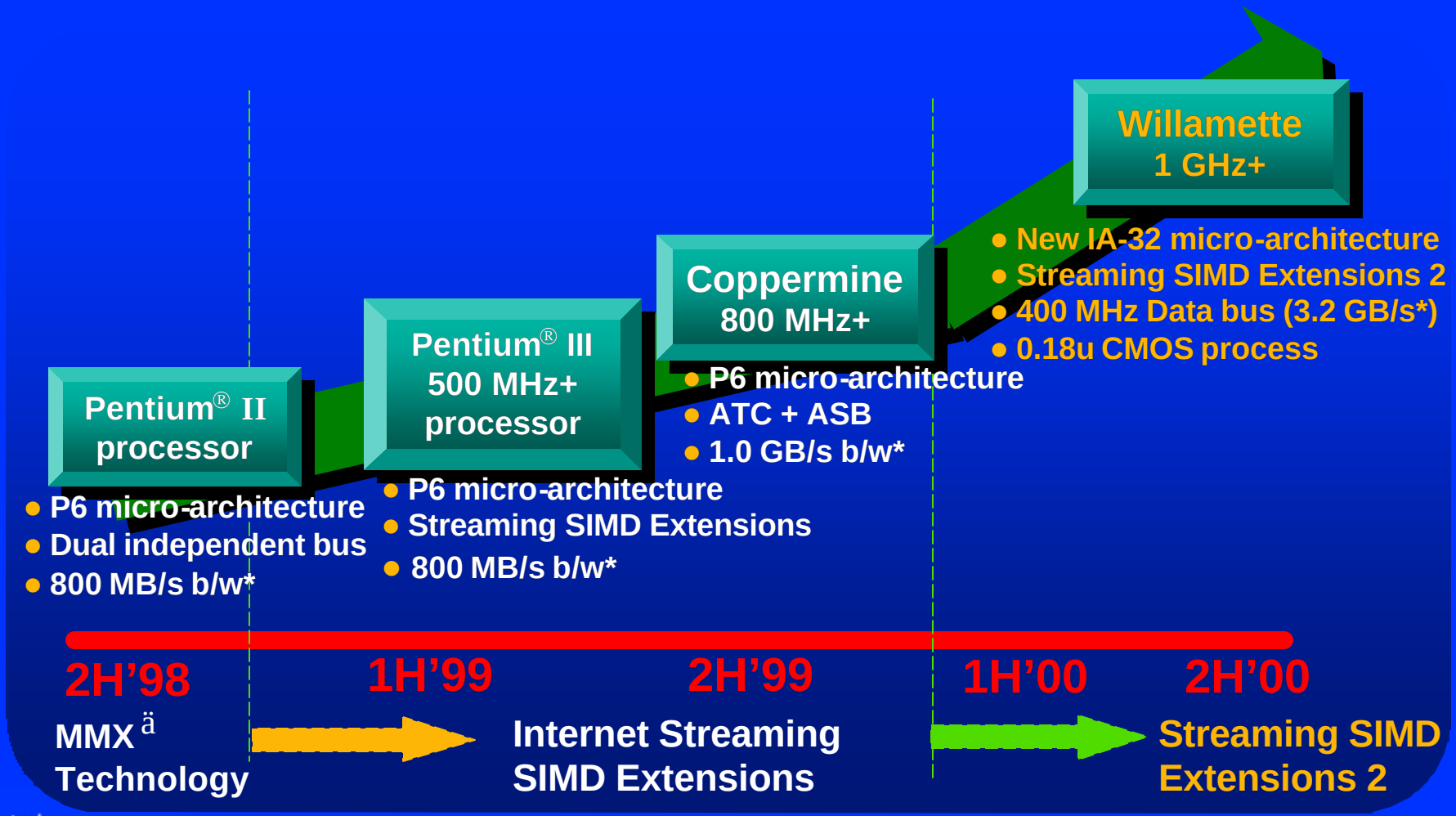
Outline

- Willamette Program Goal
- Willamette Micro-architecture
- New Instructions

Willamette Program Goal

Develop a new IA-32 micro-architecture that will drive performance leadership for the next several years.

IA-32 DT Processor Roadmap



*Note: System Bus Bandwidth.

Intel Developer Forum Spring 2000

Willamette Design

- Designed by same group as P6 core
- Built upon P6 experience
- Huge pre-silicon validation effort
  - Kept 2000-3000 CPUs busy night and day
- Booted major PC OSes within a week

Outline

- Willamette Program Goal
- Willamette Micro-architecture
- New Instructions

Willamette:
New IA-32 Micro-architecture

- Very Deeply Pipelined Core
  - *Industry leading clock rate!*
- Advanced Dynamic Execution
- Double-pumped ALU
- Execution Trace Cache
- Enhanced FP/Multi-Media
- New Instructions
- 3.2 GB/sec Bus

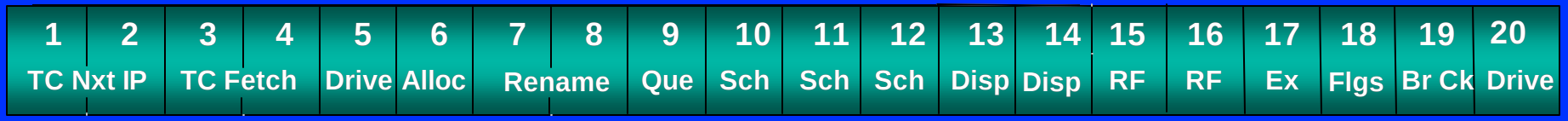
Procesor Pipelines

- Frequency depends on pipeline depth
- Example: 90 gates of logic to do
  - 30 gates per clock = 3 clocks
  - 10 gates per clock = 9 clocks
  - 10-gate version about 3x higher clock rate
- Frequency is micro-architecture decision

Willamette vs P6 Pipelines

Basic P6 Pipeline

Basic Willamette Pipeline

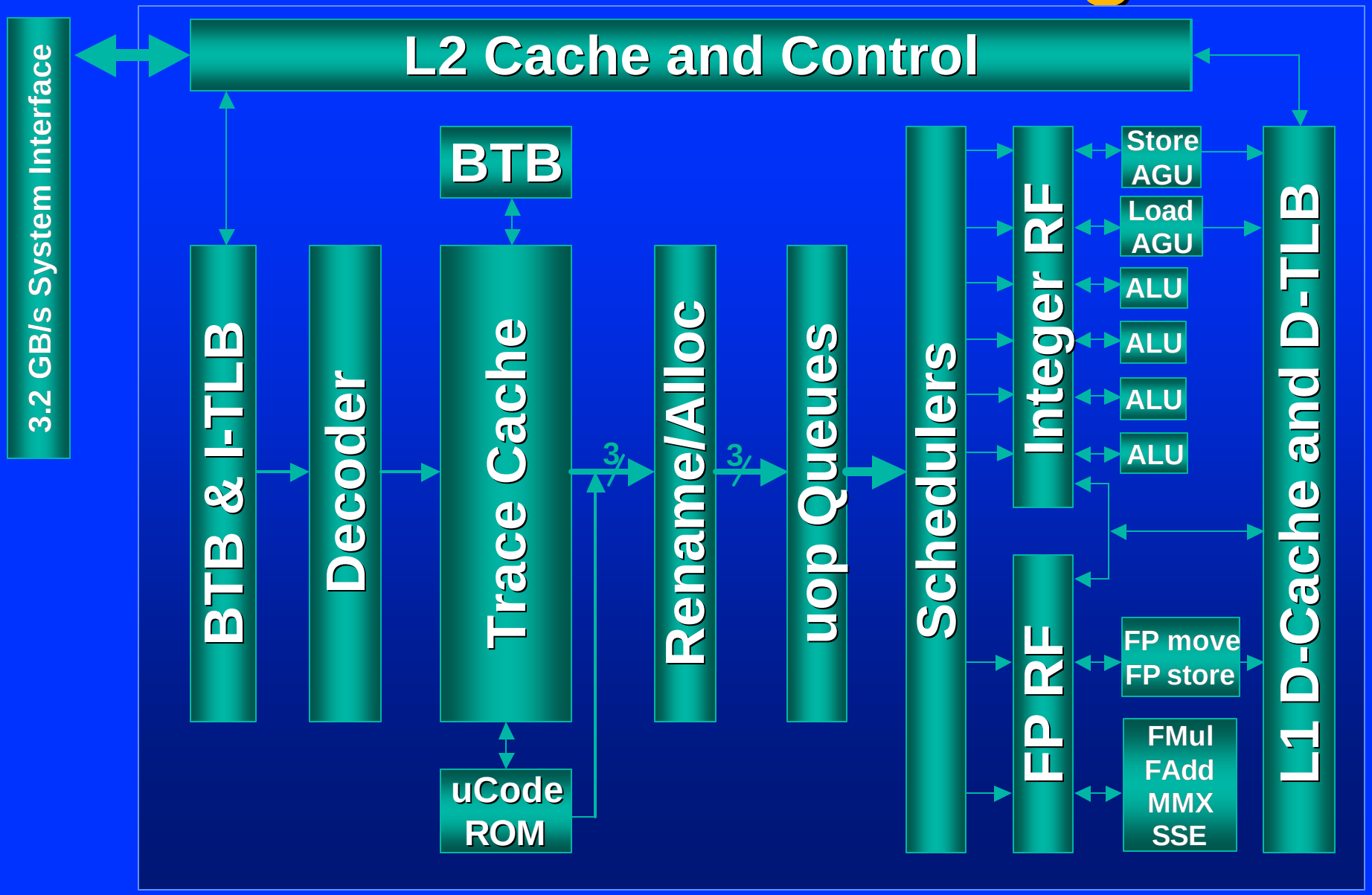


Deep pipeline enables industry leading clock rate

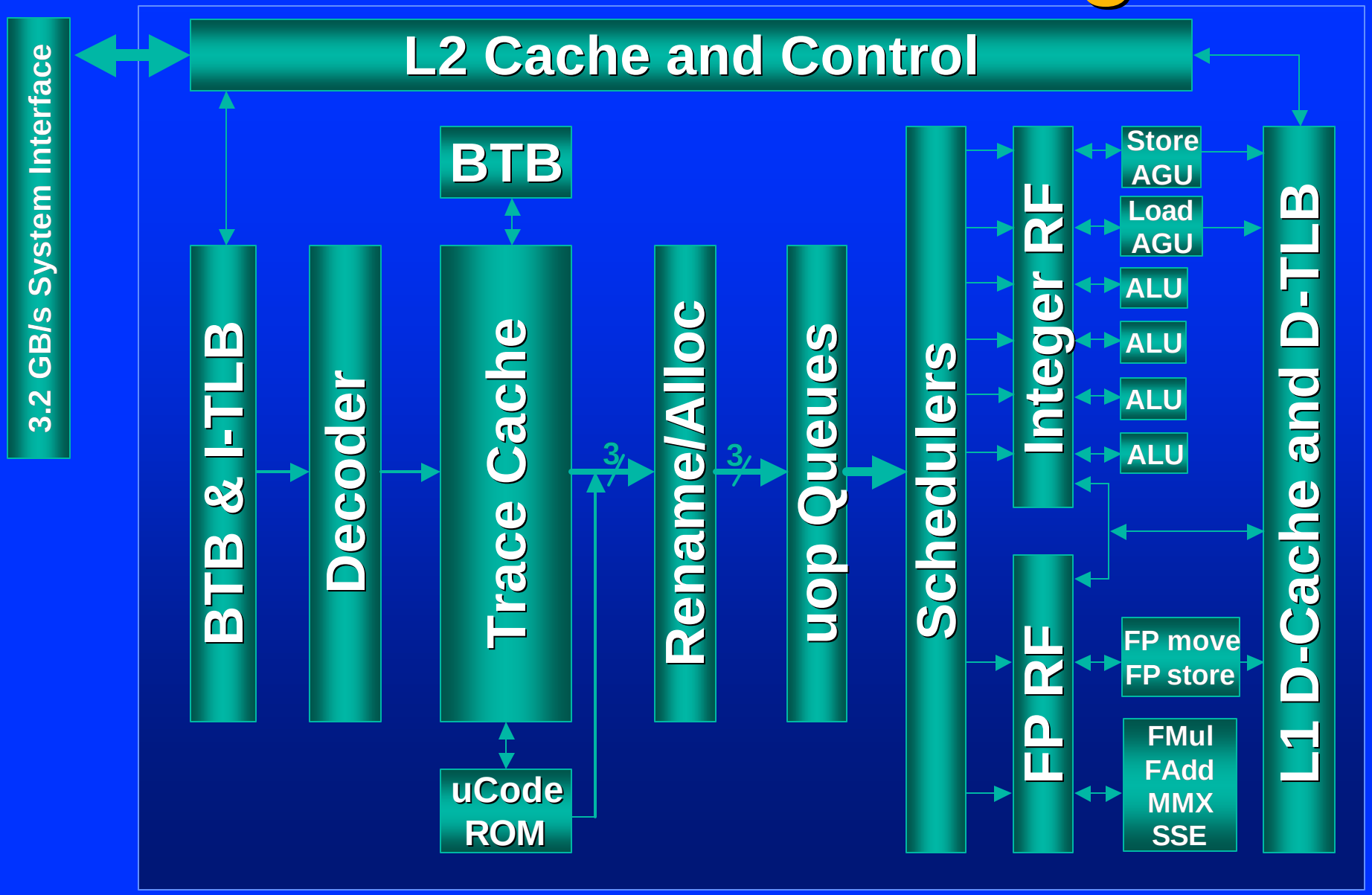
Advanced Dynamic Execution

- Extends basic features found in P6 core
- Very deep speculative execution
  - > 100 instructions in flight
  - Up to 48 loads and 24 stores in pipeline
- Improved Branch Prediction
  - Enhanced algorithm
  - 4K entry branch target array
  - Helps cover deeper pipeline
- 128 byte cache lines

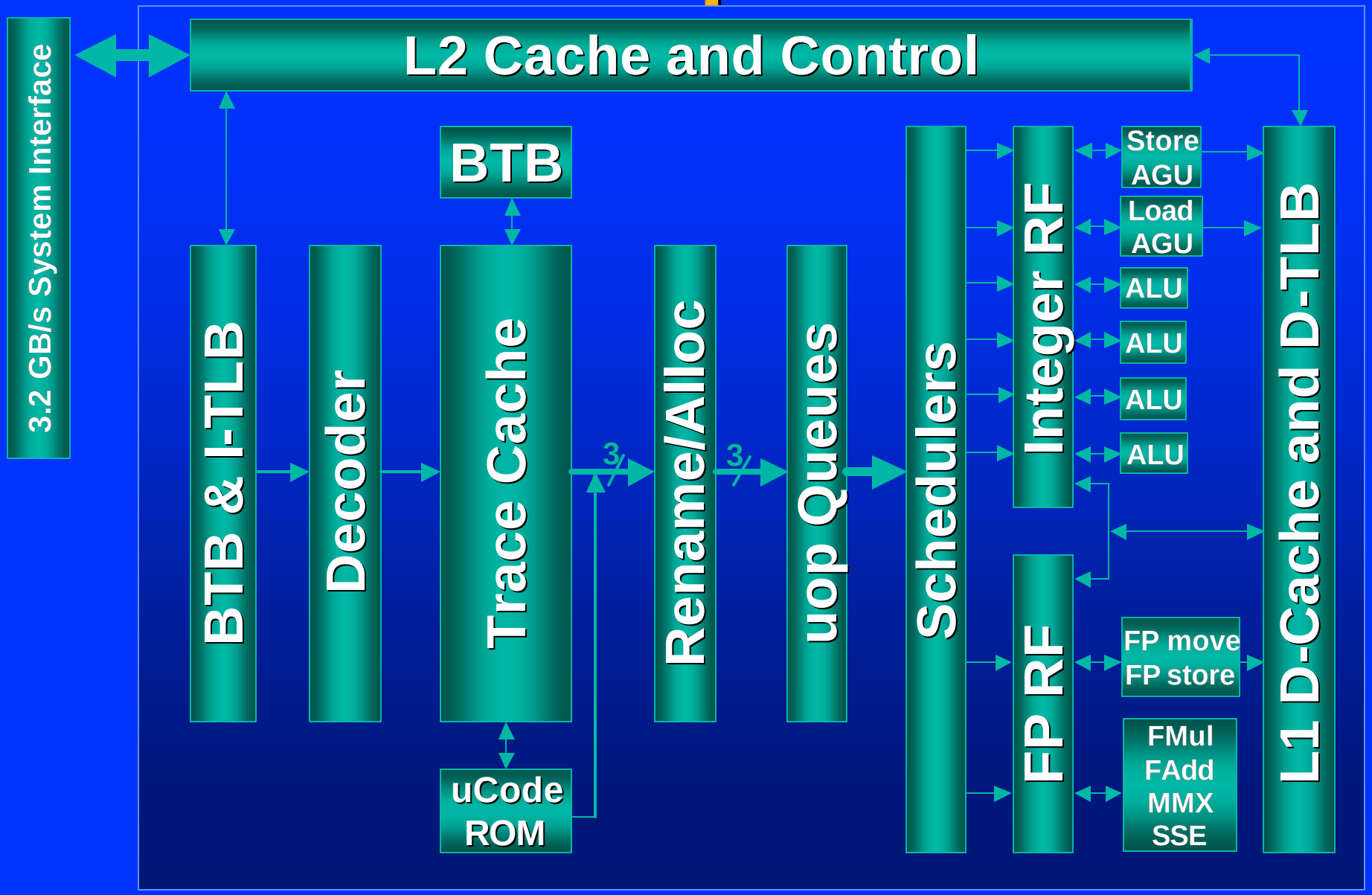
Willamette Block Diagram



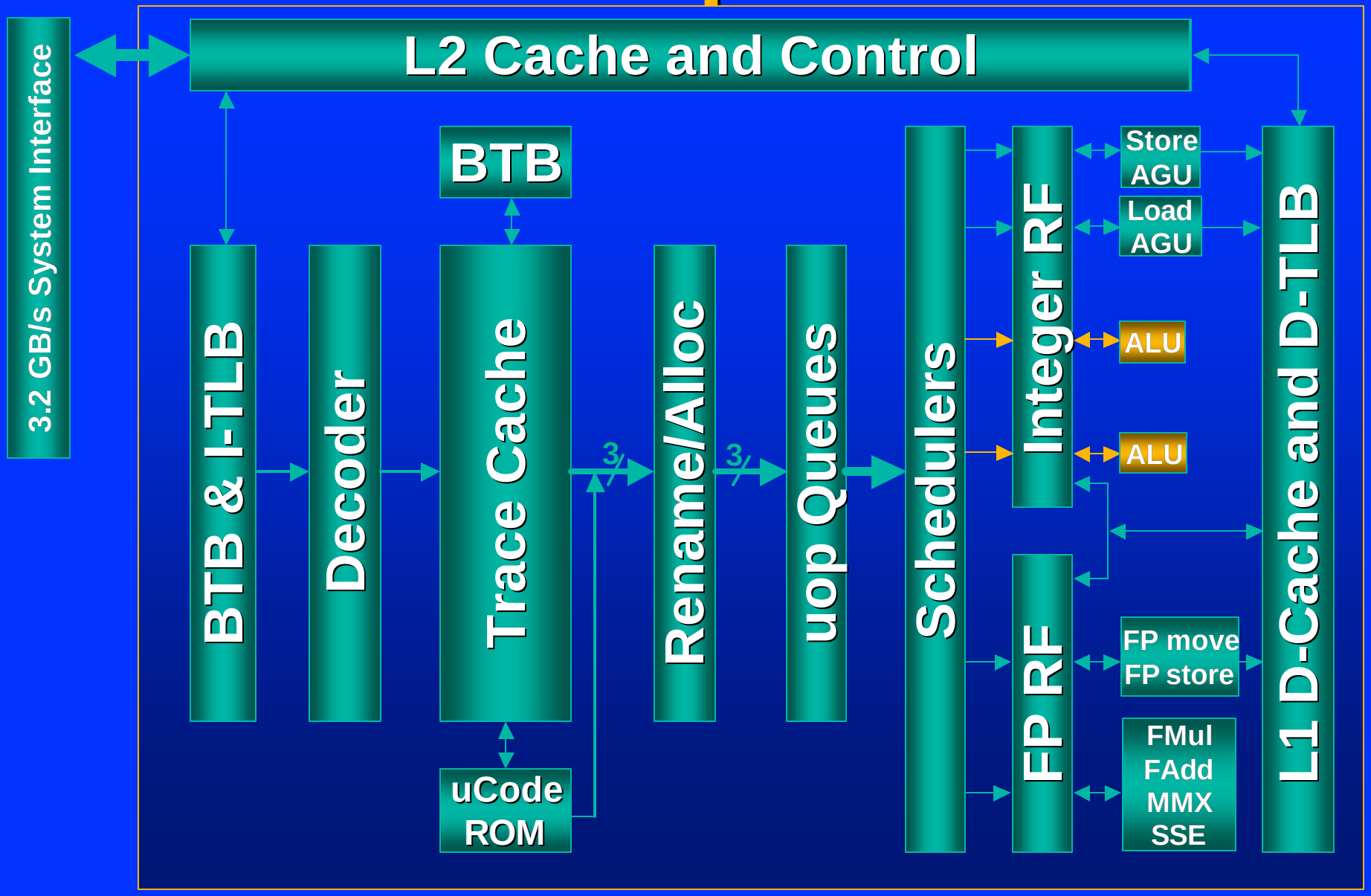
Willamette Block Diagram



Double-Pumped ALU



Double-Pumped ALU



Low Latency Execution

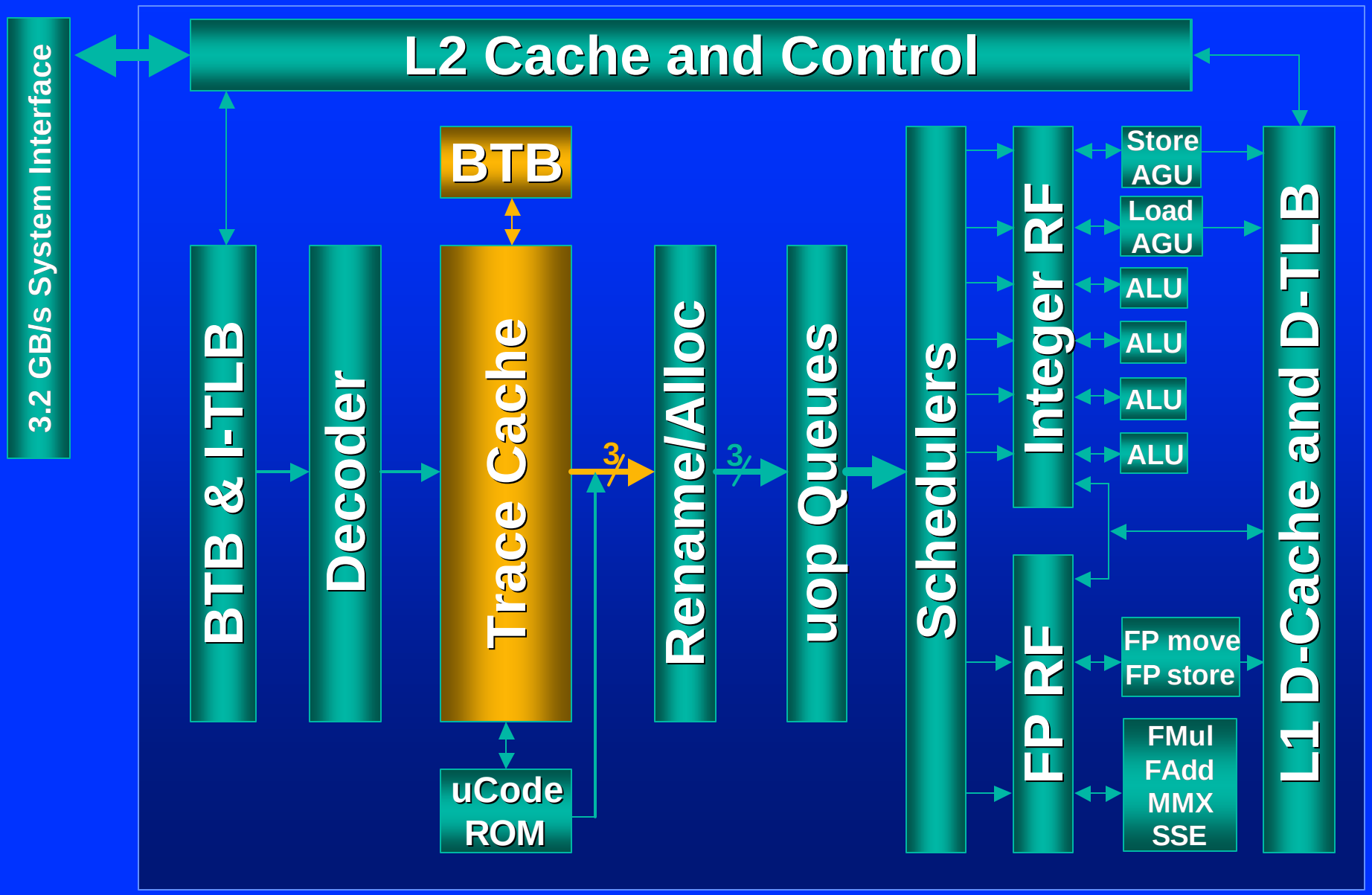
- Double-pumped Integer ALU
  - 0.5 clock latency

```
add → sub → mov → xor → …
0.5 +  0.5 + 0.5  + 0.5 = 2 clocks
```

- Low latency L1 D-cache
  - 2 clk latency

```
load → add → load → sub → …
 2   +  0.5 +   2   + 0.5 = 5 clocks
```

Execution Trace Cache

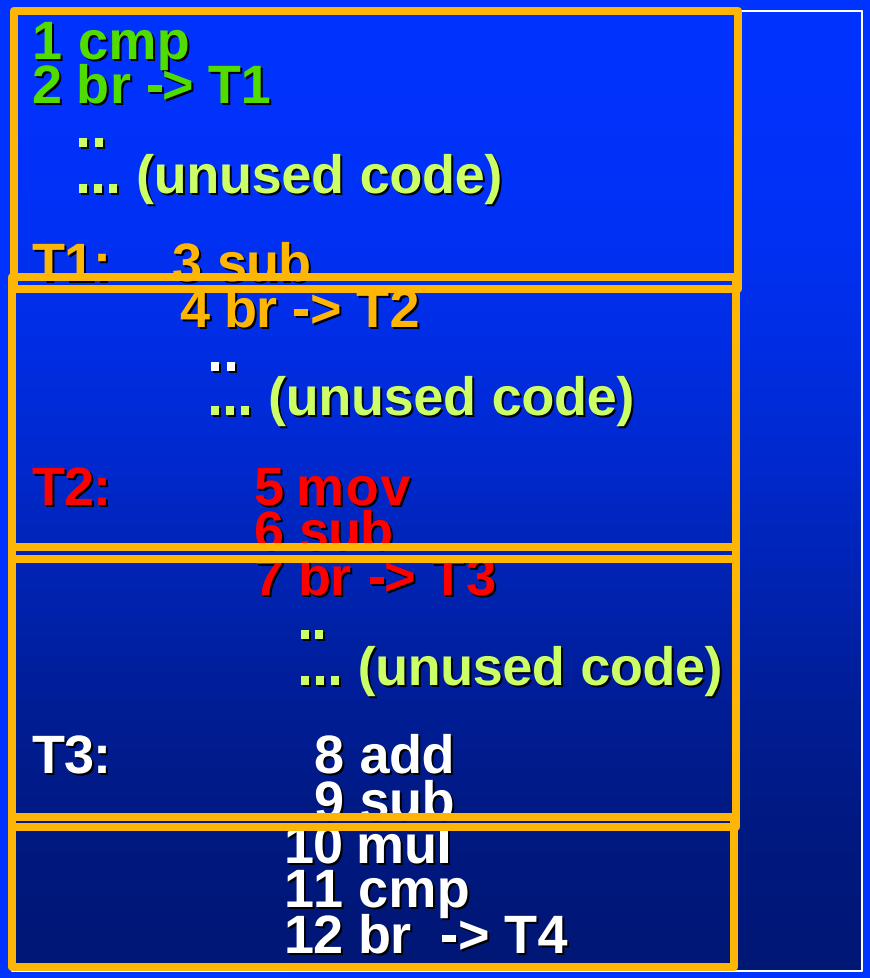


Execution Trace Cache

- Advanced form of L1 instruction cache
- Removes IA-32 decoder from main loop
  - Caches “decoded” IA-32 instructions (uops)
  - Removes decoder pipeline latency
    - Faster branch misprediction recovery
- High bandwidth IA-32 uop delivery
  - Follows predicted path of program execution
  - Integrates taken branches into single line

Execution Trace Cache feeds fast engine

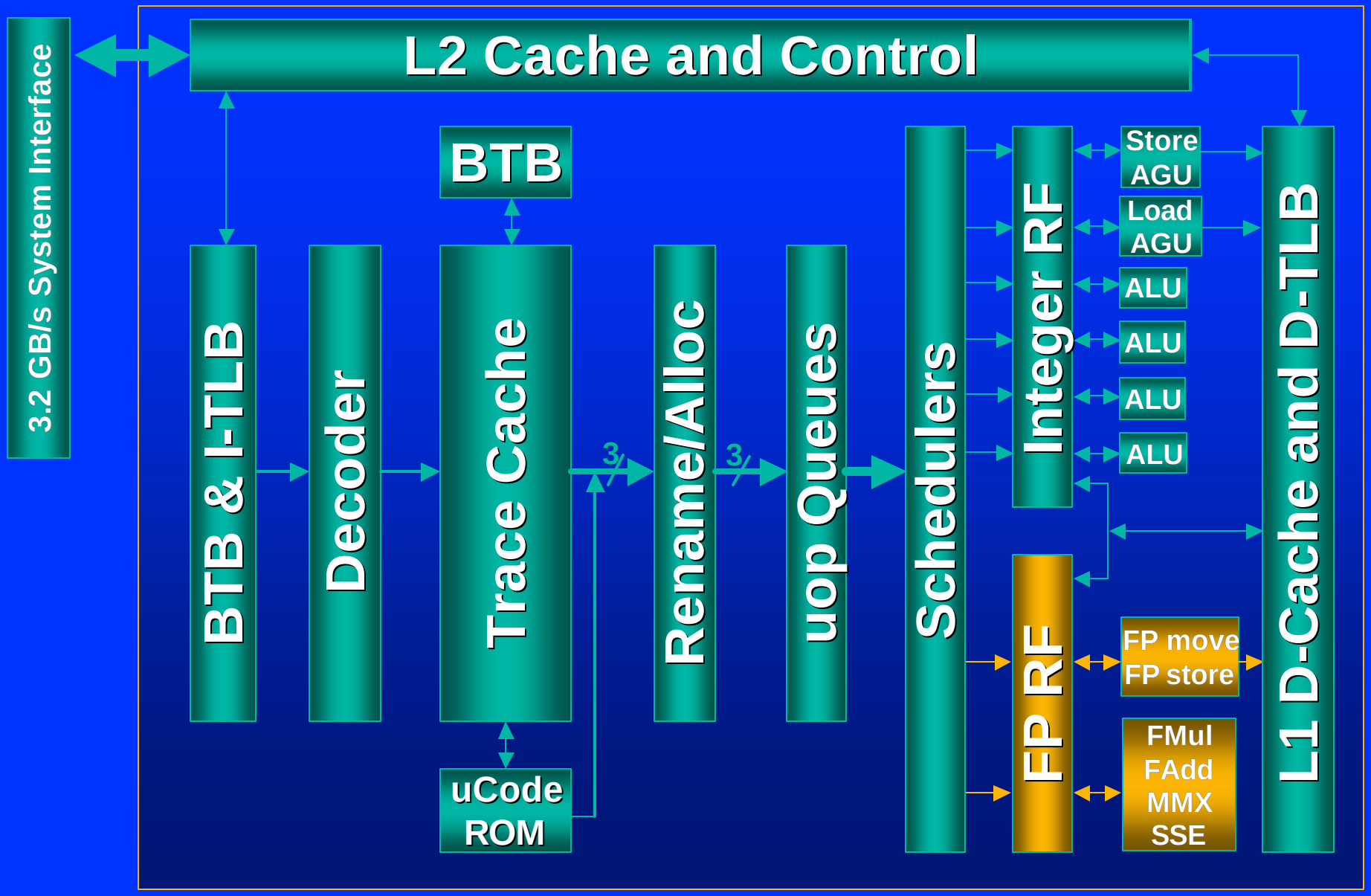
Execution Trace Cache



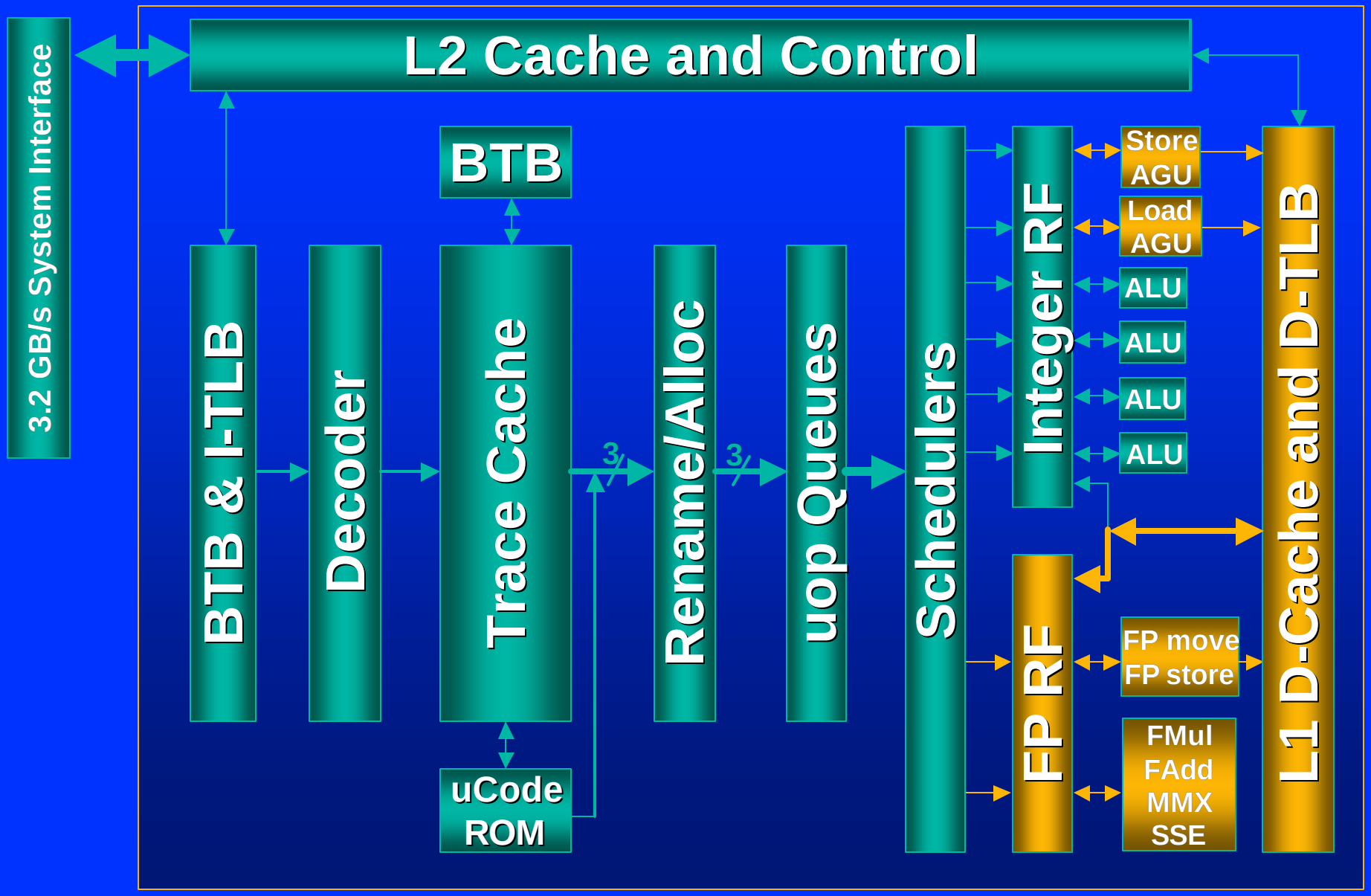
Trace Cache Delivery

1 cmp	2 br T1	3 T1: sub
4 br T2	5 mov	6 sub
7 br T3	8 T3:add	9 sub
10 mul	11 cmp	12 br T4

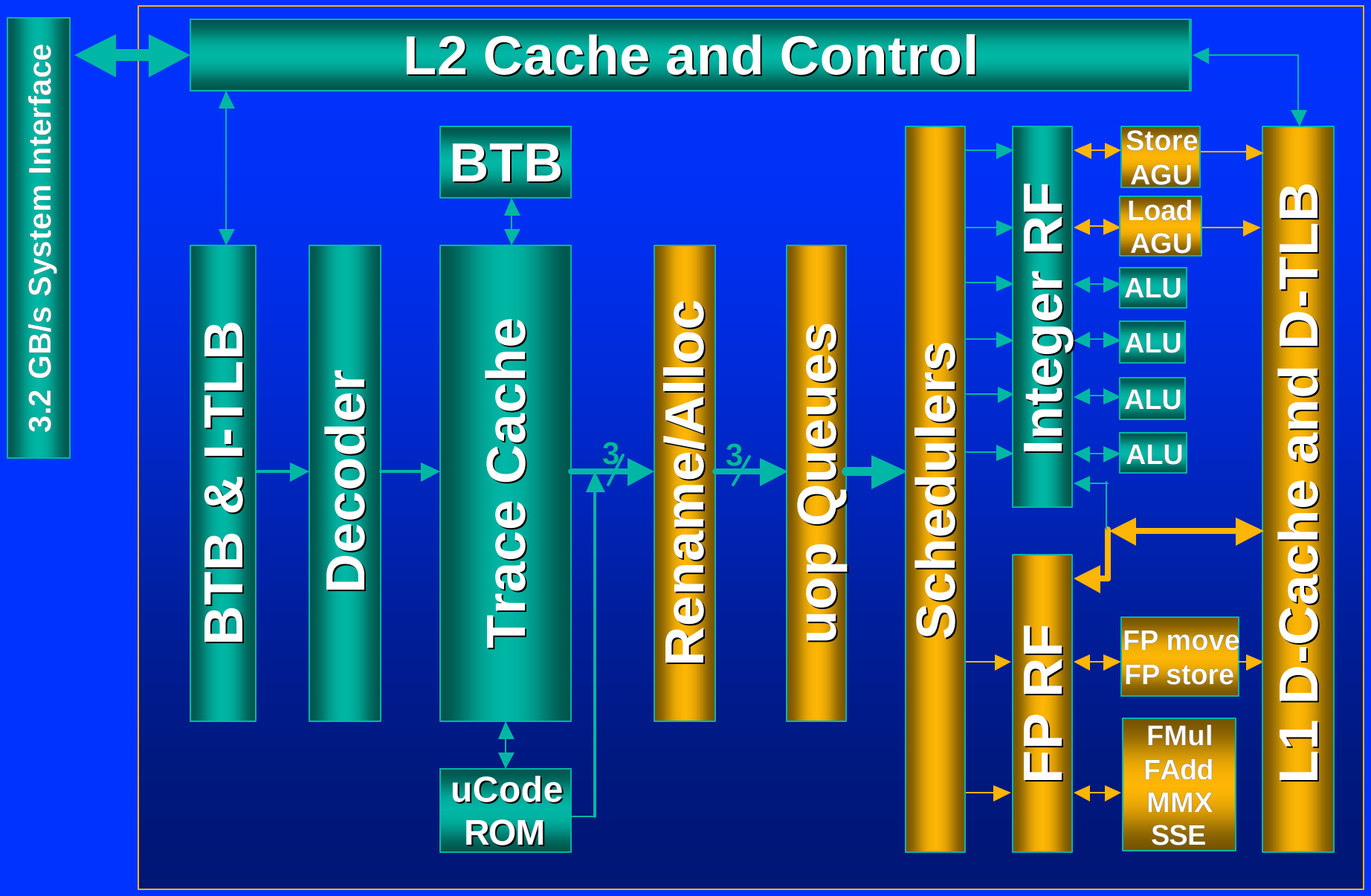
Enhanced FP/Multi-Media



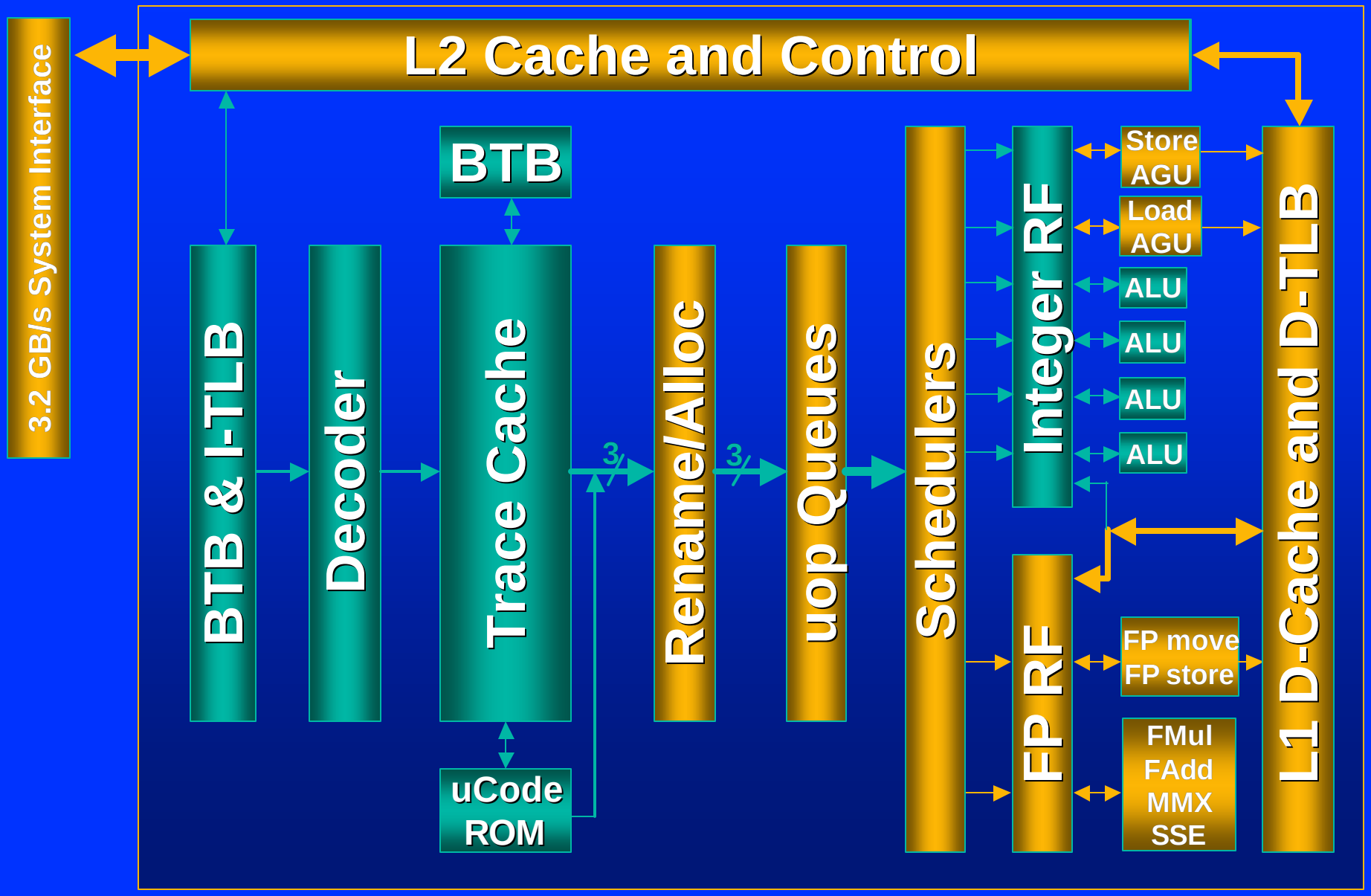
Enhanced FP/Multi-Media



Enhanced FP/Multi-Media



Enhanced FP/Multi-Media



Willamette System Bus

- 3.2 GByte/sec data transfer rate
  - 400 MHz quad pumped data bus
  - Source synchronous 64-bit data bus
- Split-transaction, deeply pipelined bus
- 128-byte lines with 64 byte accesses

High BW bus ensures performance headroom

Outline

- Willamette Program Goal
- Willamette Micro-architecture
- New Instructions

Streaming SIMD Extension 2

SSE2 Extends MMX™ and SSE technology

- **128-bit SIMD Integer arithmetic**
- **128-bit SIMD Double Precision FP**
- **Cache/memory management operations**

SSE2: 128-bit SIMD-Integer

Packed Integer (64,32,16, 8 bits)
Operations per instruction

128-bit SIMD Integer

- 64-bit MMX instructions extended to 128-bits
  - Supports the same data types
  - Easy port for most MMX applications
- Uses XMM registers instead of MMX registers
- New operations for encryption
  - Packed 32 * 32-bit unsigned multiply
  - Packed 64-bit add/subtract

Accelerates video, speech, encryption imaging/photo processing,

SSE2: 128-bit SIMD DP FP

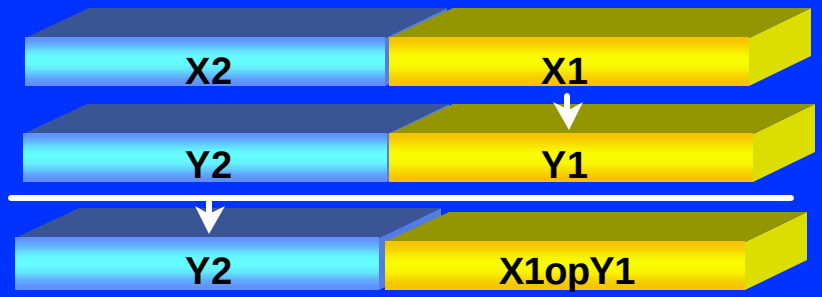
- SIMD double precision FP instructions
  - 2 x 64-bit double precision FP operations
- Full set of DP FP arithmetic ops
  - Scalar and packed DP FP operations
  - Packed/Scalar DP Û SP conversions

Accelerates Demanding Content Creation, Financial, Engineering and Scientific Apps

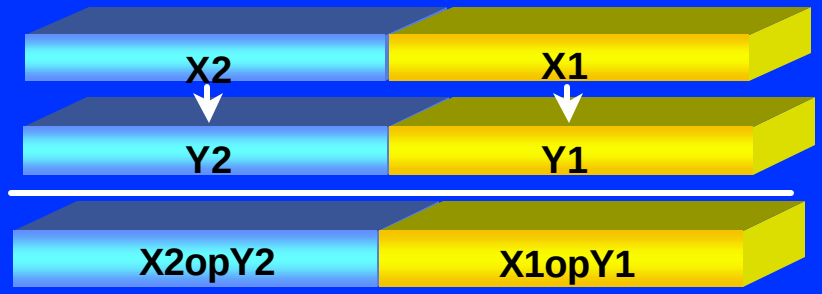
Packed & Scalar FP Operations

Double Precision FP

Scalar



Packed



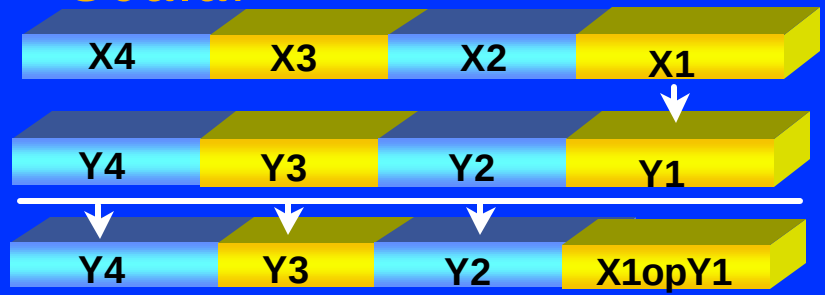
SRC

op

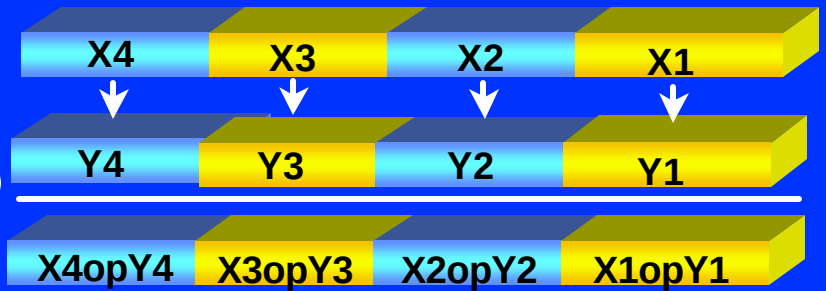
DST

Single Precision FP

Scalar



Packed



SRC

op

DST

Complete high-perf FP instruction set

SW Support for SSE2

- OS Support for SSE2
  - Windows* 98 SE
  - Windows* NT 4 with SP5 & Windows* 2000
  - Linux 2.4
- Tools
  - Intel's Willamette SDK (b in Q2'00)
  - Microsoft VC++*6.0 Service Pack
  - 3rd party s/w development tools
- Library Support
  - Intel's Performance Library
  - DirectX* 8.0, OpenGL* 1.2

Ready for Apps to use SSE2

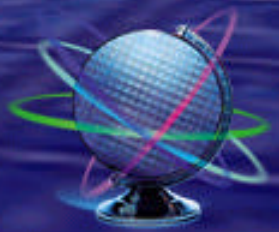
Call to Action

- ISVs
  - Attend “Optimizing Software for the Next-Generation IA-32 Architecture”
- OEMs
  - Attend “Next Generation IA-32 Desktop Platform Innovations Lab”
- http://developer.intel.com/drg
  - For Willamette program details

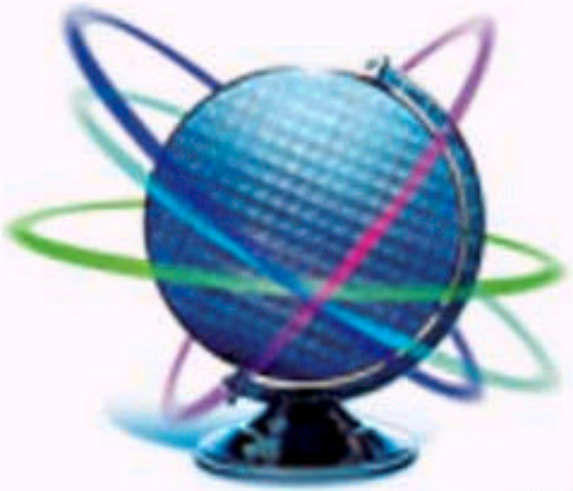
Demo

Willamette Summary

- Foundation for next generation CPUs
- Pipelined for leading edge frequency
- Delivers IA-32 performance leadership
  - Desktop apps, content creation, 3D
  - Video, imaging, speech, encryption
  - Scientific, financial, engineering



Intel
Developer
Forum
Spring 2000



intel®